STIC-ILL

NB55,187

From: Kerr, Janet
Sent: Thursday, July 12, 2001 1:20 PM
To: STIC-ILL
Subject: references for 09/457,931

Please order the following references:

Schleger et al.
Altex
18(1):5-8
2001

Prieto, P.
Science Of The Total Environment
247(2-3):349-354
2000

3728566

Genschow et al.
In Vitro & Molecular Toxicology
13(1):51-66
2000

Scholz et al.
Cells Tissues Organs
165(3-4):203-211
1999

Spielmann et al.
In Vitro Toxicology: Journal of Molecular and Cellular Toxicology
10(1):119-127
1997

Mummery et al.
Reproductive Toxicology
7(Suppl. 1):145-154
1993

Craig et al.
Biomarkers
1(2):123-135
1996

Webster, W.S.
Congenital Anom.
28(4):295-302
1989

Thanks,
Janet M. Kerr
A.U. 1633
305-4055
CM1-12A03

IN VITRO & MOLECULAR TOXICOLOGY
Volume 13, Number 1, 2000
Mary Ann Liebert, Inc.

Development of Prediction Models for Three *In Vitro* Embryotoxicity Tests in an ECVAM Validation Study

ELKE GENSCHOW,[1] GABRIELE SCHOLZ,[2] NIGEL BROWN,[3] ALDERT PIERSMA,[4] MADELEINE BRADY,[5] NICOLE CLEMANN,[6] HANNELE HUUSKONEN,[7] FRANCOISE PAILLARD,[8] SUSANNE BREMER,[9] KLAUS BECKER[10] and HORST SPIELMANN[2]

ABSTRACT

Since 1997 the National Center for Documentation and Evaluation of Alternative Methods to Animal Experiments, ZEBET, in Berlin, has been coordinating a validation study aimed at prevalidation and validation of three *in vitro* embryotoxicity tests, funded by the European Center for the Validation of Alternative Methods (ECVAM) at the Joint Research Center (JRC, Ispra, Italy). The tests use the cultivation of postimplantation rat whole embryos (WEC test), cultures of primary limb bud cells of rat embryos (micromass or, MM, test), and cultures of a pluripotent mouse embryonic stem cell line (embryonic stem cell test or EST). Each of the tests was performed in four laboratories under blind conditions. In the preliminary phase of the validation study 6 out of 20 test chemicals comprising different embryotoxic potential (*non*, *weakly*, and *strongly embryotoxic*) were tested. The results were used to define biostatistically based prediction models (PMs) to identify the embryotoxic potential of test chemicals for the WEC test and the MM test. The PMs developed with the results of the preliminary phase of the validation study (training set) will be evaluated with the results of the remaining 14 test chemicals (definitive phase) by the end of the study. In addition, the existing, improved PM (iPM) for the EST, which had been defined previously, was evaluated using the results of the preliminary phase of this study. Applying the iPM of the EST to the results of this study, in 79% of the experiments, chemicals were classified correctly according to the embryotoxic potential defined by *in vivo* testing. For the MM and the WEC test, the PMs developed during the preliminary phase of this validation study provided 81% (MM test) and 72% (WEC test) correct classifications. Because the PM of the WEC test took into account only parameters of growth

[1]Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV), Berlin, Germany.
[2]Center for Documentation and Evaluation of Alternative Methods to Animal Experiments (ZEBET) at the BgVV, Berlin, Germany.
[3]St. Georges Hospital Med. School (SGHMS), University of London, United Kingdom.
[4]Natl. Inst. of Public Health and the Environment (RIVM), Bilthoven, The Netherlands.
[5]AstraZeneca CTL, Macclesfield, United Kingdom.
[6]Novartis Pharmac AG, Basel (CH).
[7]Natl. Public Health Institute (KTL Finland), Kuopio, Finland.
[8]Synthelabo Recherche, Gargenville (F).
[9]European Center for the Validation of Alternative Methods (ECVAM), Ispra (I).
[10]Schering AG, Berlin, Germany.

and development, but not cytotoxicity data, a second PM (PM2) was developed for the WEC test by incorporating cytotoxicity data of the differentiated mouse fibroblast cell line 3T3, which was derived from the EST. This approach, which has previously never been used, resulted in an increase to 84% correct classifications in the WEC test.

INTRODUCTION

For industrial chemicals within the European Union (EU), no information is available on reproductive toxicity for the majority of base set-chemicals, which are produced at less than one ton per year or less than five tons in total. Therefore, EU regulators would appreciate better information on chemicals exhibiting reproductive toxicity early on and they are even prepared to apply *in vitro* embryotoxicity tests for regulatory purposes. However, as yet, there is no agreement on which screening test should be used. Moreover, in the field of drug development, industry is using *in vitro* tests in high throughput screening (HTS) to identify the most promising new compounds. In HTS the identification of toxic properties of new compounds has a high priority early on, even before *in vivo* testing in animals or humans is started. Because adverse effects on reproduction are among the most hazardous side effects of drugs, there is an increasing demand to include *in vitro* embryotoxicity tests in HTS systems for selecting "lead chemicals" in drug development.

For the reasons described European Center for the Validation of Alternative Methods (ECVAM) has held a workshop on screening chemicals for reproductive toxicity (Brown et al., 1995). According to the recommendations of this workshop, there is a need to validate *in vitro* tests in reproductive toxicology. Although several *in vitro* tests have been proposed as tests for developmental toxicity, none of them has successfully been validated to date. The ECVAM workshop therefore recommended that further work should be directed to improve and validate existing *in vitro* embryotoxicity tests and to introduce new ones. According to the definition of the EU, the Organization for Economic Co-operation and Development (OECD) and the U.S. NIEHS (National Institute of Environmental Health Sciences) (Balls et al., 1990, 1995; NIEHS, 1997; OECD, 1996), validation is the process by which the relevance, reliability and predictability of alternative toxicological testing methods is assessed to reduce, refine, or replace the use of laboratory animals. The main purpose of a validation study is to conduct an interlaboratory blind trial that should comprise a preliminary phase, in which a small number of coded chemicals, or "training set," is tested, and a definite phase, followed by data analysis and evaluation of the outcome of the study.

Biostatistical methods should be used throughout development and validation of alternative test methods. Hence, demonstration of reliability and relevance of an alternative method not only implicates the assessment of the variability of data, but also involves the establishment of PMs, which are able to predict the toxic effects likely to occur *in vivo* from the results obtained in the *in vitro* system.

PMs were established in the preliminary phase of the validation study of three of the most promising *in vitro* tests in a ring trial funded by the ECVAM: the rat whole embryo culture test (WEC test), the micromass test (MM test), and the embryonic stem cell test (EST). It is the goal of the ECVAM validation project on *in vitro* embryotoxicity tests to finally determine whether currently available *in vitro* methods are capable of predicting *in vivo* embryotoxic potential. Here we present the first results obtained with the PMs of the WEC and the MM test and report on the evaluation of the iPM of the EST using the results of the six test chemicals of the preliminary phase of the formal validation study conducted under blind conditions.

MATERIALS AND METHODS

Detailed standard protocols of the three *in vitro* embryotoxicity tests will be published elsewhere and will be submitted to INVITTOX, the ECVAM databank of *in vitro* techniques in toxicology.

Embryonic Stem Cell Test

The EST was performed as described previously (Scholz et al., 1999; Spielmann et al., 1997), and is an update of INVITTOX protocol no. 113 (1996). Principally, the test is composed of two procedures, a cytotoxicity test that is conducted both with the mouse embryonic stem (ES) cell line D3 (Doetschman et al., 1985) and cells of the differentiated mouse fibroblast cell line 3T3, and a differentiation assay using D3 cells. For the cytotoxicity test, cells were seeded into each well of a 96-well microtiter plate and grown in the presence of a range of concentrations of test chemical. A negative control containing solvent diluted in medium was also included. After 10 days of culture with 2 changes of medium (containing the appropriate concentration of test chemical/solvent) on Days 3 and 5, the viability of cells was determined using the MTT test (Mosmann, 1983). This test is based on the ability of mitochondrial dehydrogenases to convert the yellow substrate MTT into a dark blue formazan product, which is detected quantitatively using a microplate enzyme-linked immunoadsorbent assay (ELISA) reader. Cytotoxicity is expressed as the concentration inhibiting growth by 50% of the control level (IC_{50}3T3 and IC_{50}D3; determined from a concentration response curve).

For the differentiation assay, D3 cells were grown for 3 days in a "hanging drop" culture in the presence of a concentration range of the test chemical. During this period, D3 cells formed aggregates called embryoid bodies (EBs). After 3 days of culture EBs were transferred to bacterial petro dishes. On Day 5 EBs were plated into wells of a 24-well tissue culture plate to allow adherence and outgrowth of the EBs and development of contracting cardiac muscle cells. Differentiation was determined by microscopic inspection of EBs on Day 10. The number of wells containing contracting cardiac muscle cells was determined for each plate and compared with the number of wells containing contracting cardiac muscle cells on a solvent control plate. The inhibition of differentiation (ID_{50}) is expressed as the concentration of test chemical inhibiting the development of contracting cardiac muscle cells by 50% (calculated from a concentration response curve).

Micromass Test

The MM test is modified from the INVITTOX protocol no. 114 (1996) and uses cultures of dissociated cells of limb buds of rat embryos (Amacher et al., 1996 Flint and Orton, 1984; Kistler, 1987; Uphill et al., 1990). On Day 14 of gestation, embryos were prepared from pregnant rats. Limb buds were dissected under the microscope. After dissociation the cells were seeded into 96-well plates at high density spots ("micromass") in an appropriate culture medium, which allows differentiation into chondrocytes. After 5 days of culture in the presence of a series of concentrations of test chemicals, growth and differentiation of limb bud cells was evaluated. The extent of differentiation into chondrocytes was assessed by staining with the chondrocyte specific dye Alcian Blue (AB). The cytotoxic effect on the same cells was determined by staining with Neutral Red (NR). The concentrations inhibiting AB staining (differentiation) by 50% in comparison to unexposed controls (ID_{50}) and the concentrations reducing NR staining by 50% (IC_{50}) were estimated from concentration response curves.

Rat Whole Embryo Culture Test (WEC)

The test protocol was derived from a previous European interlaboratory validation study (Piersma et al., 1995, 1996). On Day 10 of gestation, rat embryos at the 1–5 somite stage were prepared from pregnant females. At this early stage embryos are usually more sensitive to xenobiotics than older ones. During 48 h of culture, major steps of organogenesis occur, including heart development, closure of the neural tube, development of ear and eye, branchial bars, and limb buds. Interference with development during this period may lead to general retardation of growth and development or to specific malformations in one or several organ anlagen. Embryos were cultured in rotating containers. At the end of culture, the morphology of the embryos was carefully assessed. Comparison of control embryos to exposed ones was the measure for determining the embryotoxicity of test compounds. Morphological, developmental, functional, and growth parameters were scored according to a modified scoring system first defined by Brown and Fabro (1981). Data

from seven embryos was collected for each concentration and the solvent control. The mean total morphological score (TMS; equal to the sum of all scores of the scoring system) was evaluated for each concentration in comparison to the mean TMS of the control group, which is set to 100%. In addition, number and specificity of malformations were recorded for each group of embryos. In the present report, the rate of malformed embryos of each concentration group, irrespective of the number of malformations of individual embryos was evaluated in comparison to the control group. Results were assessed by determining no effect concentrations (IC_{NOEC}), 50% effect concentrations (IC_{50}) and maximum effect concentrations (IC_{max}) from concentration response curves.

Test Chemicals

During the formal validation study, 20 chemicals and the positive control 5-Fluorouracil (5-FU) were tested under blind conditions. To provide a balanced representation of *non*, *weakly*, and *strongly embryotoxic* chemicals in the study, as well as a sufficient number of test chemicals, it was decided to test six *strongly embryotoxic*, seven *weakly embryotoxic*, and 7 *nonembryotoxic* chemicals (Table 1). The test chemicals were selected in a subcontract during the prevalidation phase of the study, on condition that the evaluated *in vivo* data and *in vitro* data from animals and humans were of high quality. The final selection and classification of test chemicals was critically reviewed and approved by the ECETOC (European Chemical Industry Ecology & Toxicology Centre, Brussels, Belgium) "Task Force on Reproductive Toxicology." The chemical selection study will be published in the final report of the formal validation study.

Coding and shipment of test chemicals was performed by RCC-Cytotest Cell Research (Roβdorf, D). The test chemicals were purchased from Merck (Darmstadt, Germany) and Sigma

TABLE 1. TWENTY SELECTED TEST CHEMICALS USED IN THE ECVAM VALIDATION STUDY[a]

Test chemical	*CAS No.*	*Company*	*Cat. No.*
	In vivo strongly embryotoxic		
5-Bromo-2′-deoxyuridine	59-14-3	Sigma	B9285
Methyl mercury chloride	115-09-3	Aldrich	442534
Hydroxyurea	127-07-1	Sigma	H8627
Methotrexate	59-05-2	Sigma	A6770
All-trans-retinoic acid	302-79-4	Roche	Dr. H. Bürgin
6-Aminonicotinamide	329-89-5	Sigma	A0630
	In vivo weakly embryotoxic		
Boric acid	10043-35-3	Sigma	B7660
Pentyl-4-yn-VPA	–	TiHo Hannover	Prof. H. Nau
Valproic acid	99-66-1	Merck	8.14439.0025
Lithium chloride	7447-41-8	Sigma	L4408
Dimethadione	695-53-4	Sigma	D7631
Methoxyacetic acid	625-45-6	Merck	8.21532
Salicylic acid sodium salt	54-21-7	Sigma	S3007
	In vivo nonembryotoxic		
Acrylamide	79-06-1	Sigma	A9099
Isobutyl-ethyl-VPA	–	TiHo Hannover	Prof. H. Nau
D-(+)-Camphor	464-49-3	Sigma	C9380
Dimethyl phthalate	131-11-3	Merck	8.00918
Diphenhydramine hydrochloride	147-24-0	Sigma	D3630
Penicillin G sodium salt	69-57-8	Sigma	PEN-NA
Saccharin sodium hydrate	82385-42-0	Sigma	S1002

[a](Smith et al., 1983; Shepard, 1992; Spielmann, 1998.)

(Deisenhofen, Germany) except the following ones: all-trans-retinoic acid (RA) was supplied by Dr. Heinrich Bürgin (Hoffmann La Roche, Basel, Switzerland), isobutyl-ethyl-VPA and pentyl-4-yn-VPA was donated by Prof. Heinz Nau (Zentrumsabteilung für Lebensmitteltoxikologie der Tierärztlichen Hochschule Hannover, Germany).

As is common practice in validation studies under blind conditions, all laboratories have to assume that each test chemicals may be severely toxic, and sensitive to temperature and UV irradiation. Thus, all chemicals had to be handled according to the most appropriate safety regulations.

Biostatistical Analysis

The statistical data analysis focused on the development of PMs for the WEC test and the MM test and for the assessment of the quality of the iPM model developed for the EST (Scholz et al., 1999). Endpoint values were determined in participating laboratories by employing EXCEL spreadsheets, which were designed for data recording of each *in vitro* embryotoxicity test. Values for IC_{50}, ID_{50}, IC_{NOEC}, and IC_{max} were calculated by estimation from concentration response curves and were determined in two independent runs of the EST and MM test, and for mean values of a total number of at least seven embryos in the WEC test. Results of individual experiments in the EST and the MM test were evaluated independently; in the WEC test, results of seven embryos were pooled and evaluated as a single experiment.

The PMs were calculated using stepwise discriminant analysis (Norušis, 1994).

RESULTS

According to the recommendations of ECVAM for the prevalidation and validation of toxicity test procedures, Standard Operating Procedures (SOPs) of the three tests were worked out in the lead laboratories during the prevalidation phase of the study and were transferred to the second laboratory (Curren et al., 1995). The SOPs for the three *in vitro* embryotoxicity tests were successfully established in the second laboratory and reproducible results were obtained. As a result of the prevalidation study, the three methods were able to discriminate a limited set of uncoded test chemicals according to their embryotoxic potential (Scholz et al., 1998).

During the formal validation study, which is currently under way, 20 chemicals are being tested under blind conditions in each of four European laboratories. In the preliminary phase of the validation study, 6 of the 20 chemicals (two chemicals not embryotoxic *in vivo*, 2 weakly embryotoxic, and 2 strongly embryotoxic) were tested.

To maintain the blind character of the study and to allow evaluation of the prediction model of the three *in vitro* embryotoxicity tests in the preliminary phase of the validation study, the embryotoxicity classes were decoded by the subcontractor for coding and shipment of the chemicals, RCC Cytotest Cell Research (Roβdorf, Germany), without providing information on the identity of the chemicals. This information was used by the biostatistician in the present study on PMs.

The remaining 14 test chemicals will be used for the evaluation of the PMs of the MM and WEC test. Because an improved PM has already been established for the EST, all 20 test chemicals will be used for the evaluation of the iPM.

Structure of the Validation Study

The structure of the validation study is illustrated in Fig. 1. The main contractor ZEBET (Center for Documentation and Evaluation of Alternative Methods to Animal Experiments at the BgVV, Berlin, Germany) is conducting the EST as lead laboratory; in addition, the European Center for the Validation of Alternative Methods (ECVAM, Ispra, Italy), Novartis (Basel, Switzerland), and Schering (Berlin, Germany) are participating with the EST. The MM test is being performed by the lead laboratory St. George's Hospital Medical School (SGHMS, University of London, UK) and the National Institute of Public Health and the Environment (RIVM, Bilthoven, Netherlands), KTL Finland (Kuopio, Finland), and Synthélabo Recherche (Gargenville, France). The WEC test is being performed by the lead laboratory RIVM (Bilthoven, Netherlands), SGHMS (London, UK), Novartis (Basel, Switzerland), and AstraZeneca CTL (Mac-

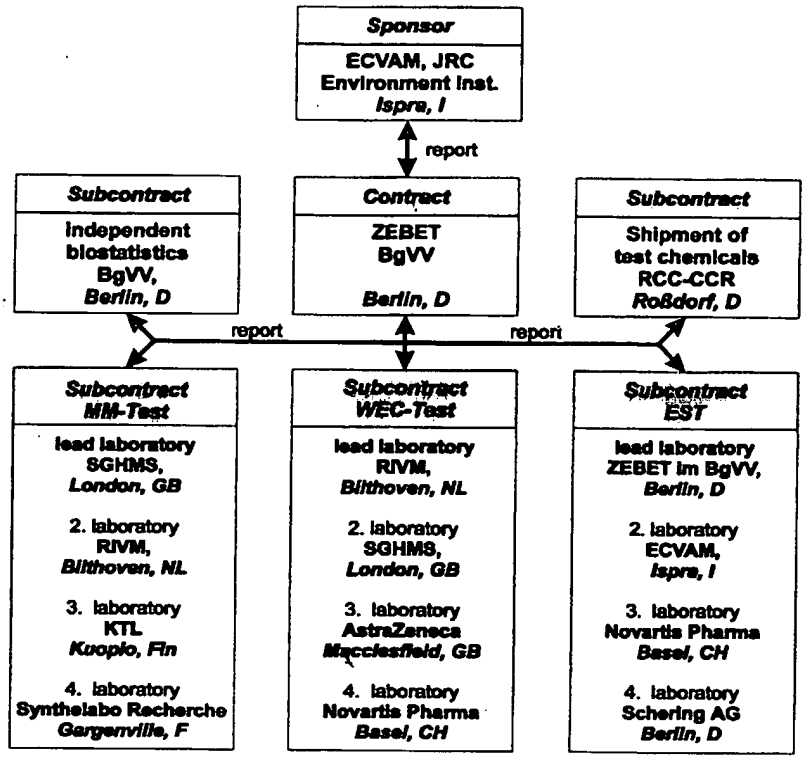

FIG. 1. Structure of the validation study. In Fig. 1 the contractor and subcontractors are shown. The subcontractors report the results to ZEBET. ZEBET reports summaries to the sponsor ECVAM. The interdependencies are marked by arrows (WEC, rat whole embryo culture test; MM, micromass test; EST, embryonic stem cell test)

clesfield, UK). The laboratories are coded A, B, C, and D for the WEC test, E, F, G, and H for the MM test, and I, J, K, and L for the EST.

Development of Prediction Models Exemplified by the EST

During a formal validation exercise, *in vitro* tests have to prove their reliability and relevance (Balls et al., 1990, 1995; NIEHS, 1997; OECD, 1996). For the first requirement, reliability, individual laboratories have to show a sufficiently high level of accuracy of independently repeated results. The second requirement, relevance, is estimated applying a biostatistically based PM. In the present study the PM should be able to discriminate between the three classes of embryotoxicity, *non*, *weakly* and *strongly embryotoxic*, as determined by *in vivo* data.

A PM is applied to use the results of an *in vitro* toxicity test to predict the toxic potential of a chemical *in vivo*. The goal of the present study was to identify variables of the PM that are important for distinguishing among the classes of embryotoxicity and to develop a procedure for correctly classifying new chemicals. For the PMs used in this study, discriminant analysis was chosen as the mathematical model. Discriminant analysis allows the determination of algorithms to distinguish among the three classes of embryotoxicity. Ideally, one variable is sufficient to distinguish among the three classes. However, in most cases two or more variables that are analysed together are necessary to distinguish between the three classes adequately (Backhaus et al., 1996). This procedure was used for the three *in vitro* embryotoxicity tests.

First, the concentration response curves from single experiments of several test chemicals were used to identify potential endpoints and variables. As an example, the endpoints and the

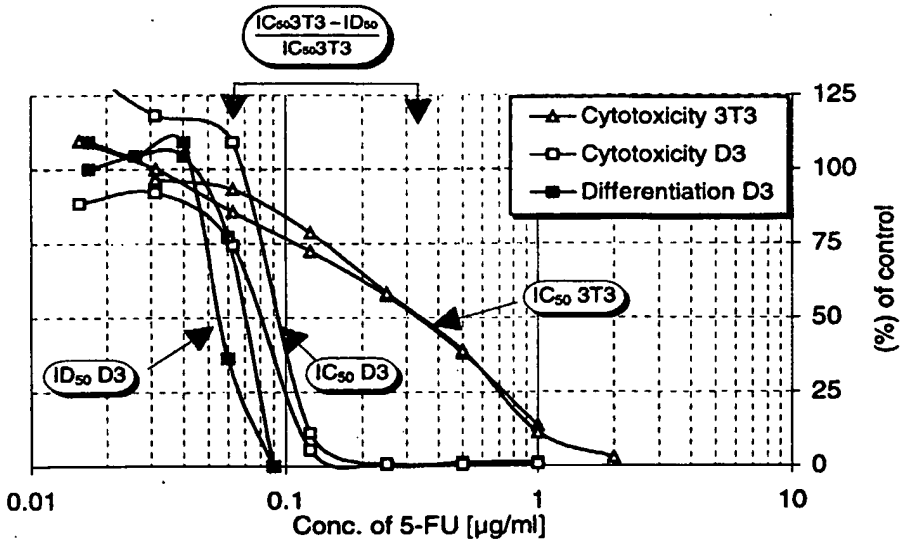

FIG. 2. Endpoints and variables used in the improved prediction model (iPM). Diagram shows concentration response curves for the three endpoints cytotoxicity of D3 and 3T3 and inhibition of differentiation of D3 cells. The curves were obtained from single experiments using the positive control chemical 5-FU. Endpoints and variables are marked by arrows.

variable applied in the improved PM (iPM) of the EST are shown in Fig. 2, which displays the results of testing the positive control chemical 5-FU.

The three endpoints ($IC_{50}3T3$; $IC_{50}D3$; and ID_{50}) cannot be used directly as variables in the linear analysis of discriminance. To convert the numerical value of the endpoint into a linear model, the logarithm of the 50% inhibition concentrations was calculated. The term "endpoint" is used only for the 50% growth inhibition concentration ($IC_{50}3T3$, $IC_{50}D3$) and the 50% differentiation inhibition concentration (ID_{50}), calculated from the concentration response curve. The term "variable" is applied to all mathematical calculations (logarithm, relative distance) of the original endpoints.

EST

Improved Prediction Model. The concentration response curves of two independent, valid experiments of the positive control chemical 5-FU are shown in Fig. 2. The endpoints and the variable are marked by arrows. To allow inclusion of occasionally divergent test results for viability and differentiation in the model, the relative distance between the two endpoints $IC_{50}3T3$ and ID_{50} was used as a variable. This variable is independent of the absolute concentration values,

TABLE 2. LINEAR DISCRIMINANT FUNCTIONS I, II, AND III OF THE EST

Function I	$5.92\ \mathrm{Ig}(IC_{50}3T3) + 3.50\ \mathrm{Ig}(IC_{50}D3) - 5.31\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 15.7$
Function II	$3.65\ \mathrm{Ig}(IC_{50}3T3) + 2.39\ \mathrm{Ig}(IC_{50}D3) - 2.03\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 6.85$
Function III	$-0.125\ \mathrm{Ig}(IC_{50}3T3) - 1.92\ \mathrm{Ig}(IC_{50}D3) + 1.50\ \frac{IC_{50}3T3 - ID_{50}}{IC_{50}3T3} - 2.67$

because this information is already given by the two other variables, the Ig(IC_{50}3T3) and the Ig(IC_{50}D3). Thus, the *in vitro* test provided a total of five different variables (Ig(IC_{50}3T3); Ig(IC_{50}D3); Ig(ID_{50}); relative distance between IC_{50}3T3 and ID_{50}; ratio of IC_{50}3T3 and ID_{50}) each of which may contribute to distinguish between classes of embryotoxicity. A stepwise selection of variables was performed using analysis of discriminance (procedure: "stepwise" of SPSS; Norusis, 1994). In a stepwise fashion, each of the variables is added separately to the model and rejected again, if it does not improve the separation of three classes of embryotoxicity significantly. Three of the variables were accepted in the analysis. For the EST, the linear discriminant functions incorporating the three variables are given in Table 2.

The following procedure was applied to precisely classify the chemicals according to the PM: The IC_{50}3T3, the IC_{50}D3, and ID_{50} were determined and used in the three linear discriminant functions. If the result of function I exceeds the results of functions II and III, the chemical is classified *nonembryotoxic*; if the result of function II exceeds the results of functions I and III, the chemical is classified *weakly embryotoxic*. Finally, if the result of function III exceeds the results of functions I and II, the chemical is classified to be *strongly embryotoxic*.

The classification of the training set of chemicals (for more details see Scholz et. al., 1999) according to the iPM resulted in 93% correct classifications (accuracy; Table 3a). Correct classifications are marked by grey boxes. The number of correctly classified individual experiments relative to the sum of all experiments is given as percent accuracy. A model usually fits the training set of chemicals better than it will fit new chemicals. Taking this limitation into consideration, the percentage of correctly classified individual experiments is an over optimistic estimation. It must therefore be expected that the accuracy of the model will decrease when completely different chemicals are evaluated.

For the evaluation of a PM the percentage of correctly classified individual experiments is not sufficient as the only index to describe the predictive power of the discriminant functions. In addition, the rate of correct classifications observed should be compared with the rate of correct classifications that can be expected just by chance. For three classes an *a priori* probability of 33% of correct classifications can be expected merely by change.

Evaluation of the EST with Six Chemicals of the Preliminary Phase. To evaluate the iPM, the six chemicals of the preliminary phase of the blind trial were classified according to the iPM. Complete sets of data were obtained from the four participating laboratories (I, J, K, and L). Results from a total of 48 individual experiments were included in the present analysis.

The iPM discriminated sufficiently between the three classes of embryotoxicity, although it

TABLE 3. CLASSIFICATION RESULTS OF THE EST (SCHOLZ ET AL., 1999) USING THE IMPROVED PREDICTION MODEL (iPM): (a) 93% OF SINGLE EXPERIMENTS OF THE TRAINING SET OF CHEMICALS AND (b) 79% OF SINGLE EXPERIMENTS IN THE VALIDATION WERE CORRECTLY CLASSIFIED

	Embryotoxicity in vivo	*No. of test chemicals*	*No. of experiments*	*Predicted class* in vitro 1	2	3
(a) Training set of chemicals	non embryotoxic[a]	3	9	9	0	0
	weakly embryotoxic[b]	3	9	1	8	0
	strongly embryotoxic[c]	4	12	0	1	11
(b) Evaluation with 6 test chemicals	non embryotoxic	2	16	10	6	0
	weakly embryotoxic	2	16	2	14	0
	strongly embryotoxic	2	16	1	1	14

[a]Class 1.
[b]Class 2.
[c]Class 3.

was developed with the results of only ten test chemicals in a prevalidation study (Table 3a and Scholz et al., 1999). With the six chemicals of the preliminary phase, 79% of individual experiments were correctly classified (Table 3b), confirming the above statement, that a decrease in the correct classification rate must be expected when chemicals different from the training set are evaluated.

MM test: Development of the Prediction Model

Laboratories E and H submitted complete data sets. From the positive control experiments of laboratory F, which usually serve as a quality control of concurrent experiments, IC_{50} values could not be computed because the curves did not cover the relevant range of concentration response according to the acceptance criteria defined in the SOP. Therefore, the data of this laboratory were not evaluated. For three of the six test chemicals tested in laboratory G, only one valid experiment was available instead of two. Including the positive control experiments, a total of 33 individual experiments were assessed in the present analysis.

In addition to the results of the preliminary phase of this trial, data from the prevalidation phase of the study were included to calculate the discriminate analysis (29 experiments of the prevalidation and 33 experiments for the 6 chemicals of the preliminary phase of the validation study).

According to the recommendations of ECVAM for the prevalidation of alternative toxicological methods (Curren et al., 1995), PMs should be derived as one of the major objectives of a prevalidation study. During the prevalidation trial of this study, only three test chemicals and a positive and a negative control chemical were tested. However, a data set of five test chemicals is considered to be insufficient for developing a PM. Furthermore, the chemicals tested during prevalidation exhibited an almost identical potential for cytotoxicity and inhibition of differentiation. To calculate a PM that is able to discriminate test chemicals characterized by divergent potential for cytotoxicity and inhibition of differentiation, and also to increase the number of test chemicals, results of the prevalidation and of the preliminary phase of the validation study were included.

The procedure for developing the PM was comparable to the procedure described above for the EST. Concentration response curves were analyzed for two endpoints, cytotoxicity determined by NR staining and inhibition of differentiation determined by AB staining. Figure 3 shows the results of a positive control experiment using 5-FU. The *in vitro* test provided a total of four experimental variables [Ig(IC_{50}), Ig(ID_{50}), relative distance between IC_{50} and ID_{50}; and ratio of IC_{50} vs. ID_{50}]; each of them may contribute to distinguish between classes of embryotoxic chemicals. Although no systematic dif-

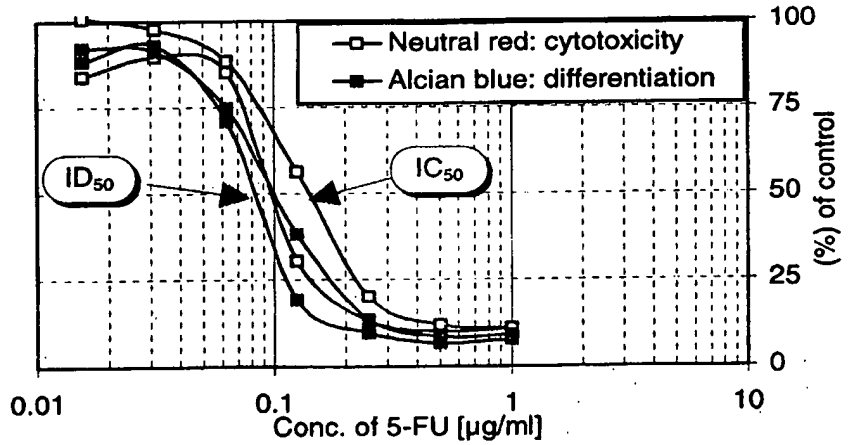

FIG. 3. Endpoints of the MM-test. Diagram shows concentration response curves for the two endpoints of the MM test: cytotoxic effect determined by staining with Neutral Red (IC_{50}) and inhibition of differentiation assessed by staining with Alcian Blue (ID_{50}). The curves were obtained from two single experiments. Endpoints are marked by arrows.

TABLE 4. LINEAR DISCRIMINANT FUNCTIONS I, II, AND III OF THE MM TEST

Function I	$6.65 * Ig(ID_{50}) - 9.49$
Function II	$6.16 * Ig(ID_{50}) - 8.29$
Function III	$-1.31 * Ig(ID_{50}) - 1.42$

ference was apparent between cytotoxicity and inhibition of differentiation, a stepwise selection of variables was performed using analysis of discriminance. As a result, the logarithm of the concentration that inhibited differentiation by 50% [$Ig(ID_{50})$] was identified as an appropriate variable in the linear discriminant analysis. Using the variable selected for the MM test, three linear discriminant functions are shown in Table 4. The procedure for classifying chemicals according to the PM is described above.

The results of the preliminary phase confirmed the observation from the prevalidation study that no significant difference was found between cytotoxicity data and inhibition of differentiation. Therefore, the twofold rule proposed as PM in previous studies and in the INVITOX protocol no. 114 (1996 and references therein), proved insufficient to predict the embryotoxic potential of test chemicals in the present study.

The training set of chemicals provided 81% correct classifications according to the PM (Table 5). The discrimination between *non* and *weakly embryotoxic* test chemicals proved to be difficult because several of the chemicals were at the borderline between *non* and *weakly embryotoxic*. Remarkably, 100% of the *strongly embryotoxic* chemicals were classified correctly in the PM. As outlined above, a model usually fits the training set of chemicals better than it will fit new test chemicals. Thus, the percentage of correctly classified chemicals in the training set is an over optimistic estimation. The evaluation of the current PM has to await the results of testing the remaining 14 test chemicals and will be conducted at the end of the validation study.

WEC test: Development of Prediction Models

Compared with the two previous *in vitro* tests, the WEC test is a laborious test because a large number of parameters have to be assessed for each embryo. Four laboratories (A, B, C, and D) provided complete data sets for the six test chemicals and the positive and negative controls. Hence, 32 experiments including positive and negative control experiments were submitted from the preliminary phase of the validation study. This data set was used to develop the PM. Data obtained during prevalidation were not used for the following reason: during the prevalidation study, the assessment of endpoints in the WEC test was slightly different from the determination of endpoints in the present study, especially with respect to malformations. Namely, during prevalidation, severely malformed embryos as well as embryos with functional defects, for example, missing heartbeat and missing circulation, were scored as "dead." Consequently, the number and specificity of malformations was not assessed properly during prevalidation and severely malformed embryos were not distinguished from embryos with defective heartbeat and circula-

TABLE 5. CLASSIFICATION RESULTS OF THE MM TEST (29 EXPERIMENTS OF THE PREVALIDATION STUDY AND 33 OF THE PRELIMINARY PHASE OF THE VALIDATION STUDY): 81% OF THE TRAINING SET OF CHEMICALS WERE CORRECTLY CLASSIFIED

	Embryotoxicity in vivo	*No. of test chemicals*	*No. of experiments*	*Predicted class* in vitro *1*	*2*	*3*
Training set of chemicals	non embryotoxic[a]	4	24	17	7	0
	weakly embryotoxic[b]	3	17	5	12	0
	strongly embryotoxic[c]	5	21	0	0	21

[a]Class 1.
[b]Class 2.
[c]Class 3.

tion. However, in the validation trial, numbers and specificities of malformations were recorded for all embryos, independent from functional parameters, for example, heartbeat and circulation. The term "dead," which was considered a poor endpoint providing insufficient information, was not used in the validation study.

Using linear discriminant analysis two PMs (PM1 and PM2) were developed for the WEC test from data of the preliminary phase of the validation study.

Prediction Model 1 (PM1). The *in vitro* test initially provided seven different experimental endpoints:

- IC_{NOEC} for malformation = the highest concentration that has no effect on the rate of malformations;

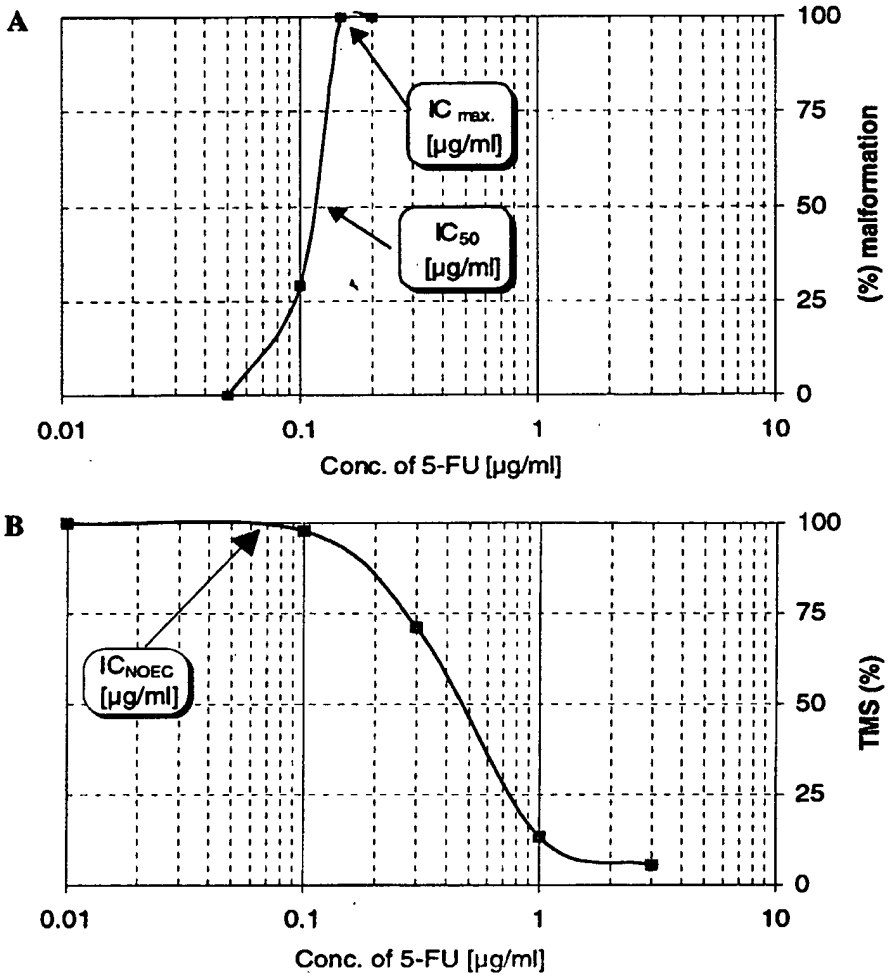

FIG. 4. Endpoints of the WEC-test. Diagram (A) shows a concentration response curve for the three endpoints for malformations of the WEC test. IC_{NOEC} is defined as the highest concentration that has no effect on the malformation rate, IC_{50} is the concentration at which 50% of the embryos are malformed and IC_{max} is defined as the lowest concentration at which a maximum malformation rate is obtained. Endpoints are marked by arrows. Diagram (B) shows a concentration response curve for the TMS (total morphological score). The endpoint IC_{NOEC} is defined as the highest concentration that has no effect on TMS (marked by an arrow).

TABLE 6. LINEAR DISCRIMINANT FUNCTIONS I, II, AND III OF PM1 OF THE WEC TEST

Function I	$18.08 * Ig(IC_{50}Mal) - 11.56 * Ig(IC_{NOEC}TMS) - 10.19$
Function II	$21.55 * Ig(IC_{50}Mal) - 15.31 * Ig(IC_{NOEC}TMS) - 10.65$
Function III	$8.70 * Ig(IC_{50}Mal) - 8.53 * Ig(IC_{NOEC}TMS) - 2.53$

TMS: total morphological score; IC_{NOEC} describes the lowest concentration, that has no effect on the TMS; IC_{50}Mal is defined as the concentration at which 50% of all tested embryos of the group are malformed. IC_{50} and IC_{NOEC} concentrations are calculated from concentration response curves by eyeball estimation.

- the highest rate of malformations observed (%);
- IC_{50} for malformations = the concentration at which 50% of the embryos are malformed;
- IC_{NOEC} for TMS = the maximum concentration that has no effect on the total morphological score (TMS);
- IC_{max} for malformations = the lowest concentration that shows a maximum rate of malformations; and
- IC_{50} for TMS and IC_{75} for TMS = the concentration at which the mean TMS is reduced to 50 or 75% of the control group of embryos;

Each of these endpoints may contribute to distinguishing among the three classes of embryotoxic chemicals. As described above for the EST, the stepwise selection of variables was performed using analysis of discriminance. Two of the variables were accepted in the analysis: IC_{50} for malformations and IC_{NOEC} for TMS (Fig. 4). Table 6 shows the PM1 calculated from data obtained in the preliminary phase of the validation study with six test chemicals and the positive and negative control. The classification criteria, which are used to discriminate between *non, weakly,* and *strongly embryotoxic* chemicals, have been described above for the EST. The classification of the training set of chemicals according to PM1 provided 72% correct classifications (Table 8). Further evaluation of the PM1 will be conducted after testing the remaining 14 test chemicals.

Prediction Model 2 (PM2). Since PM1 takes into account only parameters of differentiation and development, but not measure of cytotoxicity, we tried to improve PM1 by including cytotoxicity data provided by the cytotoxicity test with 3T3 cells obtained in the EST ($IC_{50}3T3$, Fig. 2). This approach has previously never been used as an adjunct to the WEC test. Therefore, the $IC_{50}3T3$ of the EST was incorporated in the discriminant analysis. However, this variable was not used directly in the analysis, but, in addition to the seven variables used to develop PM1, the relative distance between the $IC_{50}3T3$ and the no effect concentration of the TMS (IC_{NOEC}) of the WEC test was used as an additional endpoint. Thus, the endpoints applied in PM2 are: (1) the lowest concentration at which a maximum malformation rate is obtained (IC_{max} for malformations) and (2) the relative distance between $IC_{50}3T3$ and IC_{NOEC} for TMS (Figs. 2 and 4).

Table 7 shows the linear discriminant functions for PM2. The classification criteria, which are used

TABLE 7. LINEAR DISCRIMINANT FUNCTIONS I, II, AND III OF PM2 OF THE WEC TEST

Function I	$0.21 * \frac{IC_{50}3T3 - IC_{NOEC}\ TMS}{IC_{50}3T3} * 100 + 15.37 * Ig(IC_{max})^a - 23.58$
Function II	$0.27 * \frac{IC_{50}3T3 - IC_{NOEC}\ TMS}{IC_{50}3T3} * 100 + 17.71 * Ig(IC_{max}) - 32.37$
Function III	$0.093 * \frac{IC_{50}3T3 - IC_{NOEC}\ TMS}{IC_{50}3T3} * 100 + 4.21 * Ig(IC_{max}) - 4.23$

[a]IC_{max} refers to the lowest concentration, at which the maximum rate of malformations is observed.

TABLE 8. CLASSIFICATION RESULTS OF THE WEC TEST (32 EXPERIMENTS WITH THE FIRST SIX TEST CHEMICALS OF THE PRELIMINARY PHASE OF THE VALIDATION STUDY INCLUDING POSITIVE AND NEGATIVE CONTROL EXPERIMENT AS TRAINING SET OF CHEMICALS) (a) USING PM1 72% OF TEST CHEMICALS WERE CORRECTLY CLASSIFIED AND (b) USING PM2 84% OF TEST CHEMICALS WERE CORRECTLY CLASSIFIED

	Embryotoxicity in vivo	*No. of test chemicals*	*No. of experiments*	*Predicted class* in vitro *1*	*2*	*3*
(a) PM1	non embryotoxic[a]	3	12	7	5	0
Training set of	weakly embryotoxic[b]	2	8	2	6	0
chemicals	strongly embryotoxic[c]	3	12	0	2	10
(b) PM2	non embryotoxic	3	12	9	3	0
Training set of	weakly embryotoxic	2	8	2	6	0
chemicals	strongly embryotoxic	3	12	0	0	12

[a]Class 1.
[b]Class 2.
[c]Class 3.

to discriminate between *non*, *weakly*, and *strongly embryotoxic* chemicals, have been described above. The classification of the training set of chemicals according to PM2 provided 84% correct classifications (Table 8). Interestingly, *strongly embryotoxic* chemicals were correctly classified at a rate of 100%. It has to be stressed, however, that these are classification rates obtained with the training set, and they still have to be confirmed and evaluated by additional testing in the definitive phase of the blind trail.

DISCUSSION

In the present study we report the first results of an ECVAM validation study of three *in vitro* embryotoxicity tests, the EST, the MM test, and the WEC test. As indicated above, to date no previous validation study on *in vitro* embryotoxicity tests has been conducted under standardized conditions implemented as strictly as in the present study. The procedure implies the transfer of SOPs developed during prevalidation to secondary laboratories. Prevalidation was finished in 1998 (Scholz et al., 1998).

The validation study is being performed essentially as proposed by the ECVAM Workshop on Practical Aspects of the Validation of Toxicity Test Procedures (Balls et al., 1995), and will be finished early in the year 2000. According to recommendations of the ECVAM workshop, the main purpose of a validation study is to conduct a blind trial, as a basis for determining whether one or more tests can be shown to be relevant and reliable for the given purpose. Formal validation studies should comprise a preliminary phase, in which a small number of coded test chemicals are tested as a "training set," followed by a definitive phase, and a final evaluation of the outcome of the study. In the preliminary phase, which is the topic of the present report, a training set of six test chemicals was tested under blind conditions in the three *in vitro* embryotoxicity tests.

Despite the longtime use of established *in vitro* embryotoxicity tests (e.g., the MM test and the WEC test) for none of these *in vitro* tests has a sufficiently validated PM been developed, to allow screening of new or structurally related chemicals for embryotoxic potential in either the drug or the chemical industry. Taking into account the results of the preliminary phase, biostatistical PMs were developed for the MM test and the WEC test. This is particularly important because development of PMs is an essential prerequisite for the formal validation procedure (Curren et al., 1995; OECD, 1996).

It has to be stressed that the classification of test chemicals according to the PM in the WEC test and the MM test was based on results of the training set. These classification results provide a highly optimistic estimation, and testing with a set of new chemicals will prove the real power of the PM. This assumption is supported by our investigation of the EST. The PM of the EST had been sufficiently defined prior to the

preliminary phase of the validation study (Scholz et al., 1999), and thus underwent its first evaluation in the preliminary phase of the present study. Results obtained with the EST on the training set of chemicals in the prevalidation study provided 93% correct classifications. However, the rate of correct classifications decreased to 79%, when tested with the preliminary set of 6 test chemicals of the current validation study.

Although a wide spectrum of developmental parameters is evaluated in the WEC test (i.e., 16 parameters for general morphological development and 30 types of malformations) morphological development (total morphological score) and the percentage of malformed embryos provide sufficient information for predicting the embryotoxic potential of test chemicals in a PM. Nevertheless, the impact of the number and specificities of malformations of individual embryos determined in the WEC test will be evaluated at the end of the study.

During prevalidation of the MM test it became obvious that midbrain cells do not provide data that are reproducible in an acceptable manner. Therefore, only limb bud cells were used in the MM test in the formal validation study. Furthermore, the concentration response curves of the differentiation of limb bud cells monitored by AB staining and cytotoxicity assessed by NR uptake were almost identical. Subsequently, in the PM or the MM test only the differentiation of limb bud cells was considered because it showed a higher predictive power than the cytotoxicity to discriminate test chemicals according the three classes of embryotoxicity.

Some of the test chemicals could not easily be assigned to the groups of *weakly* and *nonembryotoxic* chemicals in all three of the *in vitro* embryotoxicity tests. This is most probably due to the limited number of test chemicals and by specific characteristics of these compounds. A comprehensive discussion of limitations of the three *in vitro* embryotoxicity tests including an evaluation of the effectiveness and accuracy of the PMs will be given by the end of the validation study.

ACKNOWLEDGMENTS

The authors are indebted to Heinz Nau (Tierärztliche Hochschule Hannover, Germany) and Heinrich Bürgin (Hoffmann La Roche, Basel, Switzerland) for supplying test chemicals and the ECETOC (European Chemical Industry Ecology & Toxicology Center, Brussels, Belgium) "Task Force on Reproductive Toxicology" for critical verification of the chemicals list. This study was supported by grant No. 12225-96-10-F1ED ISP D awarded by ECVAM at the Joint research Center (JRC) of the EU in Ispra (Italy).

REFERENCES

AMACHER, D.E., STADLER, J., SCHOMAKER, S.J., and VERSEIL, C. (1996). The comparative testing of eight coded chemicals in the rat limb bud micromass and rat embryo culture systems. ATLA 24, 945–952.

BACKHAUS, K., ERICHSON, B., PLINKE, W., and WEIBER, R. (1996). Multivariate Analysemethoden. Springer Verlag: Berlin.

BALLS, M., BLAAUBOER, B., BRUSIK, D., et al. (1990). Report and recommendations of the CAAT/ERGATT workshop on the validation of toxicity test procedures. ATLA **18,** 313–337.

BALLS, M., BLAAUBOER, B.J., FENTEM, J.H., et al. (1995). Practical aspects of the validation of toxicity test procedures. The report and recommendations ECVAM workshop 5. ATLA **23,** 129–147.

BROWN, N.A. and FABRO, S. (1981). Quantitation of rat embryonic development in vitro: A morphological scoring system. Teratology **24,** 65–78.

BROWN, N.A., SPIELMANN, H., BECHTER, et al. (1995). Screening chemicals for reproductive toxicity: The current alternatives. The report and recommendations of an ECVAM/ETS workshop (ECVAM workshop 12). ATLA **23,** 868–882.

CURREN, R.D., SOUTHEE, J.A., SPIELMANN, H., LIEBSCH, M., FENTEM, J.H., and BALLS, M. (1995). The role of prevalidation in the development, validation and acceptance of alternative methods (the first report of the ECVAM Task Force on Prevalidation). ATLA **23,** 211–217.

DOETSCHMAN, T., EISTETTER, H.R., SCHMIDT, W., and KEMLER, R. (1985). The *in vitro* development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. J. Embryol. Exp. Morphol. **87,** 27–45.

FLINT, O.P. (1993). *In vitro* tests for teratogens: desirable endpoints, test batteries and current status of the micromass teratogen test. Reprod. Toxicol. **7,** 103–111.

FLINT, O.P. and ORTON, T.C. (1984). An *in vitro* assay for teratogens with cultures of rat embryo mid-

brain and limb cells. Toxicol. Appl. Pharmacol. **76,** 383–395.

INVITTOX protocol no. 68 (1993). Embryotoxicity testing using a whole embryo culture (WEC) procedure. The ERGATT/FRAME Data Bank of In Vitro Techniques in Toxicology.

INVITTOX protocol no. 113 (1996). Embryonic Stem Cell Test (EST). The ERGATT/FRAME Data Bank of In Vitro Techniques in Toxicology.

INVITTOX protocol no. 114 (1996). *In vitro* Micromass Teratogen Assay. The ERGATT/FRAME Data Bank of In Vitro Techniques in Toxicology.

KISTLER, A. (1987). Limb bud cell cultures for estimating the teratogenic potential of compounds: Validation of the test system with retinoids. Arch. Toxicol. **60,** 403–414.

KUCERA, P., CANO, E., HONEGGER, P., SCHILTER, B., ZIJLSTRA, J.A., and SCHMID, B. (1993). Validation of whole chick embryo cultures, whole rat embryo culture and aggregating embryonic brain cell cultures using six pairs of coded compounds. Toxicol. In Vitro **7,** 785–798.

MOSMANN, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. J. Immunol. Meth. **65,** 55–63.

NIEHS (Natl. Inst. of Environmental Health Sciences, 1997). Validation and Regulatory Acceptance of Toxicological Test Methods. NIH Publication No: 97-3981. Research Triangle Park, North Carolina: NIEHS, March 1997, 105 pp.

NORUŠIS, M.J. (1994). SPSS Professional Statistics™ 6.1. Chicago, ILL: SPSS Inc. OECD (Organization for Economic Co-operation and Development, 1996). Report of the OECD Workshop on Harmonization of Validation and Acceptance Criteria for Alternative Toxicological Test Methods. OECD Publications Office, Paris, France, 60 pp.

PIERSMA, A.H., ATTENON, P., BECHTER, R., et al. (1995). Interlaboratory evaluation of embryotoxicity in the postimplantation rat embryo culture. Reprod. Toxicol. **9,** 275–280.

PIERSMA, A.H., BECHTER, R., KRAFFT, N., et al. (1996). An interlaboratory evaluation of five pairs of teratogens in postimplantation rat embryo culture. ATLA **24,** 201–209.

SCHOLZ, G., GENSCHOW, E., POHL, I., et al. (1999). Prevalidation of the Embryonic Stem Cell Test (EST), a new in vitro Embryotoxicity Test. Toxicol. In Vitro **13,** 675–681.

SCHOLZ, G., POHL, I., SEILER, A., (1998). Results of the first phase of the ECVAM project "prevalidation and validation of three in vitro embryotoxicity tests." ALTEX **15,** 3–8.

SHEPARD, T.H. (1992). *Catalog of Teratogenic Agents,* 7th ed. The Johns Hopkins University Press: Baltimore.

SMITH, M.K., KIMMEL, G.L., KOCHHAR, D.M., SHEPARD, T.H., SPIELBERG, S.P., and WILSON, J.G. (1983). A selection of candidate compounds for in vitro teratogenesis test validation. Teratogenesis Carcinog. Mutagen. **3,** 461–480.

SPIELMANN, H. (1998). Reproduction and development. Environ. Health Perspect. **106,** 571–576.

SPIELMANN, H., POHL, I., DÖRING, B., LIEBSCH, M., and MOLDENHAUER, F. (1997). The embryonic stem cell test (EST), an in vitro embryotoxicity test using two permanent mouse of cell lines: 3T3 fibroblasts and embrynonic stem cells. In Vitro Toxicol. **10,** 119–127.

UPHILL, P.F., WILKINS, S., and ALLEN, J.A. (1990). In vitro micromass teratogen test: Results from a blind trial of 25 compounds. Toxicol. In Vitro **4,** 623–626.

Address reprint requests to:
Horst Spielmann
ZEBET at the BgVV
Diedersdorfer Weg 1
D-12277 Berlin
Germany

E-mail: zebet@bgvv.de